DEPARTMENT OF HEALTH AND HUMAN SERVICES PUBLIC HEALTH SERVICE

FOOD AND DRUG ADMINISTRATION
CENTER FOR BIOLOGICS EVALUATION AND RESEARCH

Division of Biostatistics and Epidemiology (HFM-215)

Memorandum

BLA #: *98-0012*

SPONSOR: *Centocor*

DATE: *August 5, 1998*

FROM: *Teresa Neeman, Ph.D.*

THROUGH: *Peter A. Lachenbruch, Ph.D., Chief, Biostatistics Branch*

SUBJECT: *Statistical Review: Chimeric Monoclonal Antibody (cA2) to Tumor Necrosis Factor for Inflammatory Bowel Disease (Crohn's Disease)*

TO: *Dr. Barbara Matthews, Clinical Reviewer*
Division of Clinical Trial Design and Analysis (DCTDA) HFM-584

BACKGROUND

Two double-blind, placebo controlled Phase 2 studies were submitted with this application. The first study reviewed here,T20, was designed to evaluate the efficacy of cA2 in patients with fistulizing Chron's disease. The second study, T16, was a study in patients with moderate to severe active Chron's disease. Both studies were moderate in size, with approximately 25-30 patients in each treatment arm. Although the results of both studies showed a sizeable treatment effect, concerns about safety and efficacy in the chronic use of this therapy have not yet been addressed.

STUDY T20

This was a placebo controlled double blind multicenter Phase 2 study, designed to evaluate the safety and efficacy of the monoclonal antibody cA2 in Crohn's disease patients with enterocutaneous fistulae. Patients were enrolled between May 30, 1996 and October 1, 1996, with the last (week 26) evaluation on March 31, 1997. The primary objective of this study was to compare the randomized groups with respect to the closure of fistulae. A total of 94 patients from 12 study sites, 7 sites in the United States and 5 sites in Europe, were enrolled and randomized to one of 3 treatment groups: three infusions at weeks 0, 2, and 6 of either placebo, 5 mg/kg cA2 or 10 mg/kg cA2. The randomization protocol used an adaptive stratified design with site and number of fistula (1 or greater than 1) as the strata. Efficacy evaluations were performed at weeks 2, 6, 10, 14 and 18. A patient were classified as a responder if there was at least a 50% reduction from baseline in the number of draining fistulae for at least two consecutive visits 4 weeks apart. All patients responding at week 18 were followed at weeks 22 and 26 or until loss of response.

Planned Efficacy Analyses

The proposed primary analysis was to compare the three treatment groups using Mantel-Haenszel chi-squared test. This test is described in the SAS manual as Q_{MH} and is defined as

$$Q_{MH} = (n-1)\ r^2,$$

where n is the number of subjects enrolled in the trial, and r^2 is the Pearson correlation coefficient. The Pearson correlation is computed as the correlation between rows and columns after assigning equally spaced scores to the outcome and treatment variables. (e.g. no response/response are assigned 0/1, treatments are assigned 0/1/2). Under the null hypothesis, the asymptotic distribution of this statistic is chi-squared with one degree of freedom. If this test was found to be significant at the 0.05 level, Fisher's Exact Test was to be used to compare each of the treatment groups with the placebo group.

The sponsor also proposed to compare the proportion of complete responders. Among responding patients, the median time to response and the duration of response would be summarized. Additionally, for each evaluation period, the proportion of responders in each treatment group would be calculated.

The patients' global assessment score as well as the Crohn's Disease Activity Index (CDAI) score would be summarized. However, the CDAI score for 15 patients with a stoma could not be calculated, so there was some missing data.

Data from the responding patients who are followed beyond week 18 for loss of response will be summarized and used to supplement the duration of response analysis.

Efficacy Results

The sponsor's analysis of the proportion of responding patients by treatment group is summarized in the table below.

	Placebo	5 mg/kg	10 mg/kg	total
achieved primary endpoint	8 (26%)	21 (68%)	18 (56%)	47
did not achieve primary endpoint	23 (74%)	10 (32%)	14 (44%)	47
total	31 (100%)	31 (100%)	32 (100%)	94

Table 1: Primary endpoint, Sponsor's analysis, Study T20

The SAS output from PROC FREQ (at the end of this document) gave a p-value of 0.017 for the Mantel-Haenszel chi-squared test. Additionally, this reviewer confirmed the p-values reported for the two comparisons, placebo *vs.* 5 mg/kg (Fisher's Exact Test, two-sided: p=0.002) and placebo *vs.* 10 mg/kg (Fisher's Exact Test, two-sided: p=0.02).

FDA Analysis: The clinical reviewer reviewed the photographs of the fistulae at each evaluation visit for each patient. Although it was not possible to corroborate the physician's assessment of every photograph, the reviewer made the following changes to the classification of responders:

Placebo Arm:

Patient [redacted] changed from non-responder to responder. This change was not based upon the review of the photographs, but upon the observation that this patient met criteria for a responder. At baseline, this patient had one open fistula which was closed at the 10 week and 14 week evaluation.

5 mg/kg Arm:

Patient [redacted] changed from a responder to a non-responder. The clinical reviewer noted that this patients had a huge fistula that responded but clearly never closed.

Patient [redacted] changed from a responder to ineligible. From the photograph, it appeared that the patient's single fistula was closed and dry at baseline.

Patient [redacted] changed from a responder to ineligible. All three fistulae appeared to be closed and dry.

10 mg/kg Arm:

Patient [redacted]: changed from responder to ineligible. No open fistulae were seen.

Patients with no open fistulae at baseline were excluded from the FDA primary analysis, since they were not evaluable for the primary endpoint. A summary of responders by treatment group following this reclassification is displayed in the table below.

	Placebo	5 mg/kg	10 mg/kg	total
achieved primary endpoint	9 (29%)	18 (58%)	17 (55%)	44
did not achieve primary endpoint	22 (71%)	11 (42%)	14 (45%)	47
total	31 (100%)	29 (100%)	31 (100%)	91

Table 2: Primary endpoint, FDA analysis, Study T20

The comparison across the three groups using a Mantel-Haenszel chi-squared test yielded a p-value of 0.04 (see end of report for SAS output). A two-sided Fisher's Exact Test for the comparison of placebo and 5 mg/kg gave a p-value of 0.02, and the comparison of placebo and 10 mg/kg gave a p-value of 0.07. The sponsor's analyses and the FDA analyses of the primary endpoint are summarized on the next table. Although the FDA analysis is somewhat more conservative, the results are consistent with the sponsor's analyses in the rejection of the null hypothesis at the 0.05 level.

	Mantel-Haenszel chi-squared test p-value	Fisher's Exact Test placebo vs. 5 mg/kg p-value	Fisher's Exact Test placebo vs. 10 mg/kg p-value
sponsor's analyses	0.017	0.002	0.02
FDA analyses	0.043	0.02	0.07

Table 3: Summary of p-values for primary endpoint, Study T20

Exploratory Analyses/FDA

Type of Response: The sponsor planned prospectively to compare complete responders among the three groups. We were also interested to know how many of the non-responders were, in fact, partial responders, and if the responders in the different groups tended to be complete responders or partial responders. Patients with one fistula at baseline could only be complete responders or complete non-responders, so we also divided the groups between one and multiple fistulae at baseline. For this analysis, we counted the proportion of fistulae which were closed at any two consecutive visits. We reviewed the fistula listings (Volume 120), and subdivided the responder classification in each treatment group. Responders were classified as complete response or partial response, whereas non-responders were classified as some response or no response. No statistical comparisons were made. Of note, however, is that non-responders tend not to respond at all. Among the 47 non-responders, only 3 patients, one in each arm, showed any response. These patients were — placebo), with 4/8 responding fistulae, — (5 mg/kg), with 2/6 responding fistulae, and — (10 mg/kg) with 2/6 responding fistulae. In contrast, there was a greater

proportion of partial responders in the 10 mg/kg group, although this group also had more patients with a large number of fistulae at baseline. Of the patients with at least 8 fistulae at baseline, one was randomized to placebo, one was randomized to 5 mg/kg, and 3 were randomized to 10 mg/kg. The data are summarized in the table below.

			Placebo (N=31)	5 mg/kg (N=29)	10 mg/kg (N=31)
patients with 1 fistula at baseline	non-responders		11	6	9
	responders		2	8	4
patients with multiple fistulae at baseline	non-responders	no response	10	4	4
		some response	1	1	1
	responders	partial response	3	3	6
		complete response	4	7	7
all evaluable patients	non-responders	no response	21 (68%)	10 (30%)	13 (42%)
		some response	1 (3%)	1 (3%)	1 (3%)
	responders	partial response	3 (10%)	3 (10%)	6 (20%)
		complete response	6 (20%)	15 (50%)	11 (35%)

Table 4: Summary of responders by degree of response, Study T20

Duration of Response: Since patients were to be assessed at 4 week intervals, the evaluation times were entered into the database as the 2 week visit, 6 week visit, etc. Duration of response could therefore only be measured in 4 week increments. Response was defined as having at least 50% of the baseline (≥ 3 months old) fistulae closed, even if they were not the same fistulae which defined the initial response. If a patient did not have two consecutive visits in response, then their duration of response was reported as 0-4 weeks. If a patient was a responder for two consecutive visits, but was no longer a responder at the following visit, then that patient had a response time of 4-8 weeks. If a patient was a responder at the last evaluation, then the patient had an ongoing response of duration at least as long as what was observed.

Almost all patients with less than 20 week response duration stopped responding while still on study. The exception was a placebo patient, 16006, who had 4 fistulae at baseline. All of the fistulae closed at week 14, and all but one remained closed through week 26. This patient was classified at having a response of at least 12 weeks in duration. Also, all patients classified as a > 24 week response all responded at week 2 and continued to be in response at week 26. Of the patients classified as having a response between 20-24 weeks, all but four were still in response at week 26. All four patients were in the 10 mg/kg arm. A summary of the data appear in the table below.

Duration of Response	Placebo (N=31)	5 mg/kg (N=29)	10 mg/kg (N=31)
0-4 weeks (non-responders)	22	11	14
> 4-8 weeks	0	5	2
> 8-12 weeks	1	2	3
>12-16 weeks	3*	4	4
>16-20 weeks	0	3	1
>20-24 weeks	4	2	5**
> 24 weeks	1	2	2

Table 5: Duration of response by treatment group, Study T20

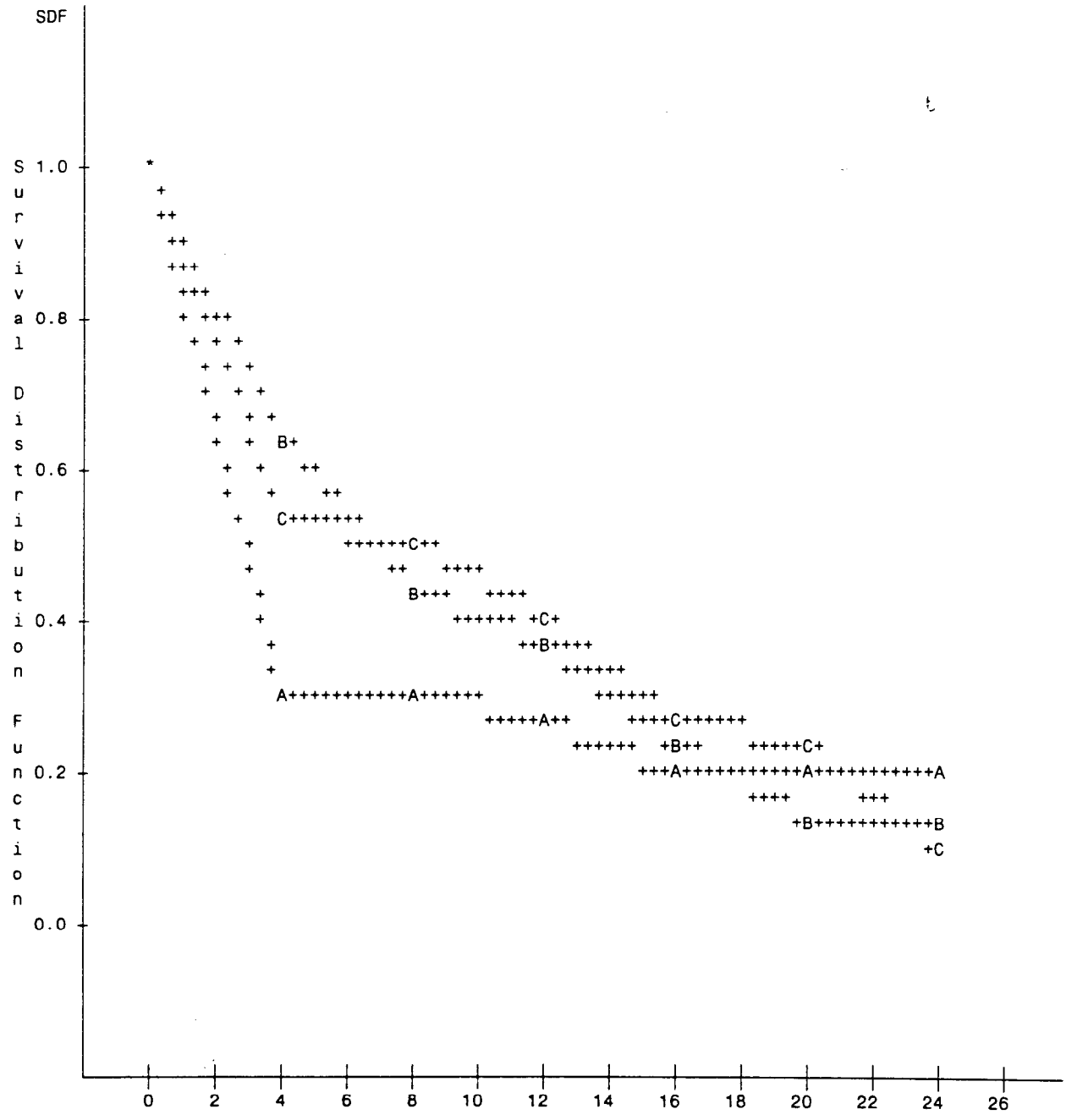
***patient 16006 had an ongoing response at week 26 after responding at week 14.**

Since these data were interval censored, they were analyzed using a life table method (PROC Lifetest, method=act). The SAS output included a graph (see below), a log-rank test comparing the three groups, and a Wilcoxon test. One can observe that since there were more responders in the treatment arms, the life table estimates for the group receiving placebo lie below those of the treatment arms. It should be emphasized that this does not imply that the responses in the placebo are were not durable, but rather reflects the lack of "true" response in this group. The life table method assumes that "failure time" is uniformly distributed across the interval, and that assumption is unlikely to hold in the 0-4 week period. Most likely, patients classified in this interval never responded, and therefore had failure time of 0.

One should also note that the life table curves converge at the 24 week point. This suggests that the responses were transitory, and there is no evidence of a lasting drug effect. Indeed, the log-rank test of the three-way comparison yielded a p-value of 0.92, suggesting that there was insufficient evidence of a difference among the three arms. The Wilcoxon test, which weights early failures more than late failures, gave a p-value of 0.17, reflecting the higher response rates in the patients treated with cA2.

The LIFETEST Procedure

Survival Function Estimates



Duration of Response by Treatment Group 13

Onset of Response: Patients in the treatment arms tended to respond earlier in the study than patients randomized to the placebo arm. The three arms were compared using a log-rank test. Patients who never responded were censored at the end of the follow-up period (Were there any drop-outs in the middle of the study? yes). The p-value for the three way comparison was 0.02, whereas the comparison of placebo *vs.* 5 mg/kg gave a p-value of 0.006. The summary of these data appear in the table below:

	Placebo (N=31)	5 mg/kg (N=29)	10 mg/kg (N=31)
week 2	3 (10%)	10 (34%)	11 (35%)
week 6	2 (6%)	5 (17%)	5 (16%)
week 10	2 (6%)	3 (10%)	0 (0%)
week 14	2 (6%)	0 (0%)	1 (3%)
never responded	22 (71%)	11 (42%)	14 (45%)

Table 6: Time to onset of response, Study T20

Treatment Effect by Investigational Site: There were 12 sites enrolling patients, six of which enrolled five patients or fewer. The largest site, #18 , had 18 patients, followed by site #22, which enrolled 16 patients. Of the 3 patients felt to be ineligible (no open fistulae at baseline), two were in site #5. No formal statistical analysis was planned, and the purpose of displaying these data was to confirm that no single study site had an undue influence in the primary analysis. In fact, in all of the centers, the patients in the combined treatment arms did uniformly better with respect to the primary endpoint than the patients in the placebo arm. The overall proportion of responders did not vary much between the centers. Among the centers with at least 8 patients, the overall proportion of success ranged from 31% (4/13) to 75% (6/8).

Study Site (# of patients)	Response Rates Placebo (N=31)	5 mg/kg (N=29)	10 mg/kg (N=31)	total response rate
18 (N=18)	2/6 (33%)	4/6 (67%)	4/6 (67%)	10/18 (56%)
22 (N=16)	1/5 (20%)	4/5 (80%)	2/6 (33%)	7/16 (44%)
2 (N=13)	0/4 (0%)	2/4 (50%)	2/5 (40%)	4/13 (31%)
5 (N=12)	2/4 (50%)	1/3 (33%)	4/5 (80%)	7/12 (58%)
3 (N=9)	1/3 (33%)	2/3 (67%)	3/3 (100%)	6/9 (67%)
16 (N=8)	2/3 (67%)	3/3 (100%)	1/2 (50%)	6/8 (75%)
1 (N=5)	0/2 (0%)	1/2 (50%)	0/1 (0%)	1/5 (20%)
6 (N=3)	0/1 (0%)	0/1 (0%)	0/1 (0%)	0/3 (0%)
12 (N=2)	0/1 (0%)	0/1 (0%)	-	0/2 (0%)
20 (N=2)	0/1 (0%)	-	1/1 (100%)	1/2 (50%)
21 (N=2)	0/1 (0%)	1/1 (100%)	-	1/2 (50%)
14 (N=1)	-	-	0/1 (0%)	0/1 (0%)

Table 7: Summary of Response by Study Site, Study T20

Baseline Factors and Treatment by Factor Interactions: We considered gender, race, baseline oral corticosteroid use , baseline azathioprine or 6 MP use, disease duration, age, and baseline CDAI score as possible covariates which may predict outcome. In the table below, the proportion of responders in each subgroup is tabulated by treatment group. In the fourth column, one can compare the overall response rate in each subgroup. Although there were small differences in the proportions of responders between subgroups, there were no statistically significant differences noted for the any of the subgroups considered below. There were, however, apparent differences in the treatment effect between men and women. Overall, there was a higher placebo response rate in women, as well as a lower response rate among women in both of the treated arms. The fifth column of this table compares the treatment effect of cA2, by combining the response rates from the 5 mg/kg arm and the 10 mg/kg arm. The odds ratio, a measure of treatment effect in each of the subgroups, is displayed in this column. When the odds ratios are very different, as can be seen between men and women, one can test (Breslow-Day or Zelen's Test in StatXact) how likely the differences could have occurred by chance. The p-value of 0.04 from the Breslow-Day test suggests that, at least for the patients on the study, these differences may be real. No other treatment by subgroup interactions were observed in this study.

Proportion of Responders by Treatment Group for Various Subgroups

Subgroup		Placebo	5 mg/kg	10 mg/kg	total	odds ratio placebo/combined
gender	male	4/17 (24%)	10/14 (71%)	9/12 (75%)	23/43 (54%)	12.7*

female	5/14 (36%)	8/15 (53%)	8/19 (45%)	21/48 (44%)	1.6
oral corticosteroid use	no	7/20 (35%)	12/18 (67%)	11/21 (52%)	30/59 (51%)	2.7
	yes	2/11 (18%)	6/11 (54%)	6/10 (60%)	14/32 (44%)	6.0
race	Afro-Amer.	1/2 (50%)	1/2 (50%)	2/3 (67%)	4/7 (57%)	1.5
	Caucasian	8/29 (28%)	17/27 (63%)	15/28 (54%)	40/84 (48%)	9.5
azathioprine or 6 MP use	no	5/22 (23%)	13/18(72%)	8/15 (53%)	26/55 (47%)	6.0
	yes	4/9 (44%)	5/11 (45%)	9/16 (56%)	18/36 (50%)	1.3
disease duration**						
	< 11 years	4/15 (27%)	9/13 (69%)	7/15 (47%)	20/43 (47%)	2.8
	≥ 11 years	5/16 (31%)	9/16 (56%)	10/16 (62%)	24/48 (50%)	3.2
age†	< 35 years	5/13 (38%)	5/9 (56%)	11/19 (58%)	21/40 (52%)	2.1
	≥ 35 years	4/18 (22%)	13/20 (65%)	6/12 (50%)	23/50 (46%)	5.1
baseline CDAI[A]						
	≤ 150	3/10 (30%)	11/12 (92%)	5/11 (46%)	19/33 (58%)	5.3
	151-220	3/6 (50%)	2/5 (40%)	1/3 (33%)	6/14 (43%)	0.6
	> 220	1/9 (11%)	3/9 (33%)	8/12 (67%)	12/30 (40%)	8.8

Table 8: Baseline factors and interactions with treatment, Study T20

*Breslow-Day Test for homogeneity of treatment effect, p=0.038, suggests there may be a stronger treatment effect among men.

** also analyzed this covariate as a continuous variable in a logistic regression model, p>0.5.

† age also analyzed as a continuous covariate in a logistic regression model, p =0.35.

[A] 15 cases dropped due to missing values.

Fistulae open more than 2 Years: It was hypothesized that fistulae of long standing duration, open for at least 2 years before the patient entered the study, may be more resistant to closing. We also considered the possibility that the treatment effect may be more pronounced in these patients. For this analysis, we considered only the 42 patients with at least one long standing fistula at baseline. A patient in this subset was classified as a responder if at least 50 % of these fistulae were not draining for at least 2 consecutive visits. Perhaps because fistulae tend to respond as all or none, every patient classified as a responder in this analysis was also a responder in the primary analysis. Likewise, there were no responders in this analysis who were non-responders in the primary analysis. However, there was no evidence that these fistulae were any more resistant to closing than the "younger" fistulae. The overall proportions of responders in the whole study were 29%, 58%, and 55% in the placebo arm, 5 mg/kg arm, and 10 mg/kg arm respectively, whereas the proportions of responders with respect to long standing fistulae in this subgroup were 31%, 72% and 55%. A summary of these data appear in the table below.

	Placebo	5 mg/kg	10 mg/kg	total
≥ 50% long-standing fistulae closed	5 (31%)	8 (72%)	8 (53%)	21 (50%)
< 50% long-standing	11 (69%)	3 (28%)	7 (47%)	21 (50%)

fistulae closed				
total	16 (100%)	11 (100%)	15 (100%)	42

Table 9: Response rates in patients with fistulae open for least 2 years, Study T20

Abdominal Fistulae: There were 9 patients in this study presenting with abdominal fistulae at baseline. In addition, one placebo patient, —— with only perianal fistulae at basline,developed an abdominal fistula at week 26 of the study. Although the numbers in this subset are too small to make any definitive conclusions, a trend of a treatment effect for this group was also apparent. A summary of the responding and non-responding patients by treatment group appears in the table below.

	Placebo	5 mg/kg	10 mg/kg	total
achieved primary endpoint	0	3	1	4
failed to achieve primary endpoint	2	1	2	5

Table 10: Response rates in patients with abdominal fistulae, Study T20

CONCLUSIONS

Although the differences in response rates between the placebo group and the cA2 treated groups were statistically significant, questions remain about the durability of response. Patients received doses at weeks 2, 4 and 6, but this dosing strategy should be thought of as one-time dosing. After 6 months of follow-up, the drug effect had disappeared and the proportion of responding patients in the placebo arm was similar to the proportions in the treatment arms. The data suggest, therefore, that although this agent has an initial beneficial effect on Crohn's disease, a single set of doses is unlikely to provide durable benefit in this chronic disease. There are no data to assess chronic use of cA2 for this indication. There is no information regarding the formation of neutralizing antibodies (HACA) with repeated dosing and how this may effect the efficacy of this product. There is also no safety data to allay concerns of a possible increase in malignancies or serious infections. The agency should carefully weigh the observed early benefits seen with this product against the paucity of information regarding the safety and efficacy of repeated use for this chronic indication.

Study T16

This study was a Phase 2, placebo controlled, multicenter trial, designed to explore the efficacy of a single infusion of cA2, and

There were two phases to this trial. In the first phase, patients were randomized to receive a single infusion of either placebo or 1 of three doses of cA2: 5 mg/kg, 10 mg/kg or 20 mg/kg. A total of 108 patients were enrolled in this first phase across 18 sites between June 21, 1995 and October 31, 1995. Of the 18 sites, 12 were in the US, accounting for 75 patients, 5 were in Europe (31 patients), and one was in [redacted] patients). The primary endpoint was the proportion of responders at week 4, a response being defined as a ≥70 point reduction in the CDAI score. Secondary endpoints were the proportion of patients with a ≥ 100 point reduction in the CDAI score, and the proportion of patients in remission at 4 weeks.

Patients responding at week 4 were followed until week 8. Those patients still responding at week 8 were eligible to enroll in the second phase of the trial

ANALYSIS PLAN

Since this was a Phase 2 trial, the statistical analysis plan was incomplete. For the primary analysis (proportion of responders in the initial phase), an overall treatment difference was to be analyzed using a chi-squared test. If this test was significant at the 0.05 level, then each treatment group was to be compared against placebo.

When analyzing change of CDAI scores from baseline, non-responding patients who received open-label cA2 would have the 4 week score carried forward for the 8 week and 12 week evaluation. It was not stated how group comparisons would be made for this endpoint.

The statistical plan for the r phase was stated as follows: "Data collected from the phase will be summarized by treatment group and analyzed for treatment group differences. The previous dose received in the initial treatment phase will be used as a covariate."

PRIMARY ENDPOINT RESULTS/ SPONSOR

The sponsor cited one patient in the placebo arm who was not evaluable for the primary endpoint. A summary of these data for the evaluable patients, together with the p-values, is presented in the table below:

	Placebo (n=24)	5 mg/kg (n=27)	10 mg/kg (n=28)	20 mg/kg (n=28)	total (n=107)
# patients achieving primary endpoint	4 (17%)	22 (81%)	14 (50%)	18 (64%)	58 (54%)
# patients who failed to achieve primary endpoint	20 (83%)	5 (19%)	14 (50%)	10 (36%)	49 (46%)
overall treatment effect (chi-squared, 3 d.f.) $p < 0.001$					
2-sided Fisher's Exact Test:	-	$p < 0.001$	p=0.02	$p < 0.001$	-

Table 11: Response rates by treatment group, Study T16

PRIMARY ENDPOINT ANALYSIS/ FDA:

Three patients in the placebo arm [redacted] and one patient in the 5 mg/kg [redacted] did not have complete CDAI evaluations at week 4, and a CDAI score was not calculated. Patients [redacted] discontinued the study before the week 4 evaluation. Patient [redacted] received an open-label 10 mg/kg infusion, but did not respond and discontinued at week 8. There were, therefore 104 patients evaluable for the primary endpoint. A summary of the numbers of responding/non-responding/unevaluable patients by treatment group is given in the table below:

	Placebo (n=25)	5 mg/kg (n=27)	10 mg/kg (n=28)	20 mg/kg (n=28)	total (n=108)
# patients achieving primary endpoint	4 (16%)	22 (81%)	14 (50%)	18 (64%)	58 (54%)
# patients who failed to achieve primary endpoint	18 (72%)	4 (15%)	14 (50%)	10 (36%)	46 (43%)
# patients not evaluable at week 4	3 (12%)	1 (4%)	0 (0%)	0 (0%)	4 (4%)

Table 12: Response rates by treatment group, FDA analysis, Study T16

We considered various analyses to account for the missing values. If the missing data points are missing at random, then the analysis based upon the 104 evaluable patients would be unbiased. On the other hand, if the patient scores were missing for reasons related to the CDAI score, then other analyses must be considered. After perusing the patient summaries, this reviewer feels it is likely to be the case that the scores were missing because these patients had not improved on therapy. If all patients with missing values are considered to be failures, then the resulting p-values are very much like the original sponsor's analysis. However, as a worst-case scenario, one can bias the data as much as possible against the treatment, and consider all patients with missing

values in the placebo arm as treatment successes and all other patients with missing values failures. Even under this situation, there is still evidence of a treatment effect. The results of the chi-squared test for overall treatment effect and individual treatment comparisons with placebo are shown in the table below:

	χ^2 test of overall treatment effect	2-sided Fisher's Exact Test		
		placebo vs. 5 mg/kg	placebo vs. 10 mg/kg	placebo vs. 20 mg/kg
evaluable (104 patients) missing = excluded	< 0.001	< 0.001	0.036	0.002
ITT (108 patients) missing = failures	< 0.001	< 0.001	0.01	< 0.001
ITT (108 patients) worst-case scenario missing(plac.)=succ. Missing(trt.)=failure	0.001	< 0.001	0.16	0.013

Table 13: Summary of p-value for analyses of primary endpoint, Study T16

Treatment Effect by Center: Of the 18 investigational sites, 13 had 5 of fewer patients enrolled. Because a patient could be randomized to one of four possible arms, these centers, considered separately provide very little information of treatment effect. The other five centers, three in the US and two in Europe, accounted for 61 of the 108 enrolled patients. The sites, 2, 3, 5, 18 and 22, were also the five largest sites in the T20 fistulae study. The overall response rates in these centers ranged from 22% (2/9) to 69% (9/13). Although the numbers of patients randomized to each arm is small, each center showed a higher response rate in cA2 treated patients. Moreover, the response rates in this subpopulation is very similar to the response rates in the population as a whole. A breakdown of clinical response at week 4 by these sites is given in the table below:

Study Site		placebo	5 mg/kg cA2	10 mg/kg cA2	20 mg/kg cA2	total response rate
2	response	2	4	3	3	12/20 (60%)
(n=20)	no response	3	0	2	2	
	data missing	0	1	0	0	
22	response	0	3	3	3	9/13 (69%)
(n=13)	no response	2	0	1	0	
	data missing	1	0	0	0	

Study Site		placebo	5 mg/kg cA2	10 mg/kg cA2	20 mg/kg cA2	total response rate
3	response	0	1	0	1	2/9 (22%)
(n=11)	no response	1	2	2	2	
	data missing	2	0	0	0	
18	response	0	1	0	3	4/9 (44%)
(n=9)	no response	3	1	1	0	
	data missing	0	0	0	0	
5	response	1	2	1	0	4/8 (50%)
(n=8)	no response	1	0	1	2	
	data missing	0	0	0	0	
total response rate		3/16 (19%)	11/15 (73%)	7/14 (50%)	10/16 (62%)	31/61 (51%)

Table 14: Response Rates for the largest study sites, Study T16

Other Endpoints: For these exploratory analyses, we included all patients treated (n=108), where patients with no CDAI scores at 4 weeks were considered to be treatment failures. In other clinical trials in Crohn's disease, ≥ 100 point reduction from baseline was used as a measure of clinical response. The table below shows the number of patients achieving a ≥ 100 point reduction from baseline at the 4 week evaluation. Although a treatment effect is still evident (chi-squared test- p=0.01), the size of the effect is less, and the effect appears to be more uniform across the cA2 treatment groups.

	Placebo (n=25)	5 mg/kg (n=27)	10 mg/kg (n=28)	20 mg/kg (n=28)	total (n=108)
# patients with ≥ 100 point reduction at week 4	4 (16%)	15 (56%)	12 (43%)	16 (57%)	47 (44%)
# patients with < 100 point reduction at week 4	21 (84%)	12 (44%)	16 (57%)	12 (43%)	61 (56%)
overall treatment effect (chi-squared, 3 d.f.) p = 0.0097					
2-sided Fisher's Exact Test:	-	p = 0.004	p = 0.04	p = 0.004	-

Table 15: 100 point reduction in CDAI scores, by treatment group, Study T16

In order to get a more sensitive comparison of the distribution in change of CDAI scores among groups, we plotted the histograms of the change scores for each of the treatment groups. As can

be seen in the figure below, there were no unusual outliers in any of the treatment groups. There was, therefore, no evidence in this population that cA2 exacerbated disease.

In addition, the remission at week 4 and compared summary of these table below. Since CDAI scores above patients in clinical 4 were also primary endpoint. in remission at week remission through patients were no at week 12; 5 in the where most of the observed and 2 in arm.

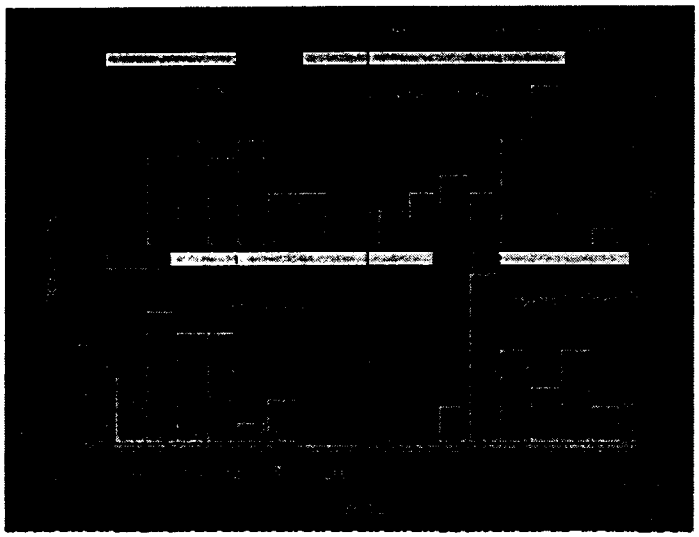


Figure 1: Reduction in CDAI Scores from baseline to week 4

patients in were summarized between groups. A data appears in the all patients had 220 at baseline, all remission at week responders by the Of the 28 patients 4, 21 remained in week 12. Seven longer in remission mg/kg cA2 arm, remissions were the 10 mg/kg cA2

	Placebo (n=25)	5 mg/kg (n=27)	10 mg/kg (n=28)	20 mg/kg (n=28)	total (n=108)
# patients in clinical remission at week 4	1 (4%)	13 (48%)	7 (25%)	7 (25%)	28 (26%)
# patients not in clinical remission at week 4	24 (96%)	14 (52%)	21 (75%)	21 (75%)	80 (74%)
overall treatment effect (chi-squared, 3 d.f.) p = 0.004					
2-sided Fisher's Exact Test:	-	p < 0.001	p = 0.05	p = 0.05	-

Table 16: Proportion of patients in clinical remission at week 4, Study T16

THIS SECTION
WAS
DETERMINED
TO BE NOT
RELEASABLE